Hand-related physical function in rheumatic hand conditions

Klokker, Louise; Terwee, Caroline; Wæhrens, Eva Elisabet Ejlersen; Henriksen, Marius ; Nolte, Sandra; Liegl, Gregor; Kloppenburg, Margaret; Westhoven, Rene; Wittoek, Ruth; Kjeken, Ingvild; Haugen, Ida K; Schalet, Ben; Gerschon, Richard; Bliddal, Henning ; Christensen, Robin

Published in:
B M J Open

DOI:
10.1136/bmjopen-2016-011174

Publication date:
2016

Document version
Final published version

BMJ Open

Hand-related physical function in rheumatic hand conditions: a protocol for developing a patient-reported outcome measurement instrument

Louise Klokker,[1] Caroline B Terwee,[2,3] Eva Ejlersen Wæhrens,[1,4] Marius Henriksen,[1,5] Sandra Nolte,[6,7] Gregor Liegl,[6] Margreet Kloppenburg,[8,9] Rene Westhoven,[10] Ruth Wittoek,[11] Ingvild Kjeken,[12] Ida K Haugen,[13] Ben Schalet,[14] Richard Gershon,[14] Henning Bliddal,[1] Robin Christensen[1]

To cite: Klokker L, Terwee CB, Wæhrens EE, *et al*. Hand-related physical function in rheumatic hand conditions: a protocol for developing a patient-reported outcome measurement instrument. *BMJ Open* 2016;**6**:e011174. doi:10.1136/bmjopen-2016-011174

▸ Prepublication history for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2016-011174).

Received 15 January 2016
Revised 11 October 2016
Accepted 27 October 2016

For numbered affiliations see end of article.

Correspondence to
Louise Klokker; Louise.Klokker.Madsen@regionh.dk

ABSTRACT

Introduction: There is no consensus about what constitutes the most appropriate patient-reported outcome measurement (PROM) instrument for measuring physical function in patients with rheumatic hand conditions. Existing instruments lack psychometric testing and vary in feasibility and their psychometric qualities. We aim to develop a PROM instrument to assess hand-related physical function in rheumatic hand conditions.

Methods and analysis: We will perform a systematic search to identify existing PROMs to rheumatic hand conditions, and select items relevant for hand-related physical function as well as those items from the Patient Reported Outcomes Measurement Information System (PROMIS) Physical Function (PF) item bank that are relevant to patients with rheumatic hand conditions. Selection will be based on consensus among reviewers. Content validity of selected items will be established through the use of focus groups. If patients deem necessary, we will develop new items based on the patients' input. We will examine whether it is valid to score all selected and developed items on the same scale as the original items from the PROMIS PF item bank. Our analyses will follow the methods used for calibrating the original PROMIS PF item bank in US samples, which were largely based on the general PROMIS approach.

Ethics and dissemination: This study will be carried out in accordance with the Helsinki Declaration. Ethics approvals will be obtained where necessary, and signed informed consent will be obtained from all participants. We aim to disseminate the results of the study through publication in international peer-reviewed journals and at international conferences.

Strengths and limitations of this study

- The aim points to fulfil a need for a new patient-reported outcome measurement (PROM) to assess hand function.
- The proposed methodology follows the best practice standards for developing PROMs.
- The use of Computer Adaptive Testing (CAT) potentially reduces patient burden in terms of time to fill out questionnaires.
- The domain of the outcome measurement instrument is limited to function, based on the assumption that function is a pivotal construct that also relates to other aspects of a health condition (eg, pain).The patient population is limited to osteoarthritis and rheumatoid arthritis with hand involvement.

INTRODUCTION

Physical function is a concept covering the 'ability to carry out activities that require physical actions ranging from self-care (activities of daily living) to more complex activities that necessitate a combination of skills, often within a social context' (http://www.nihpromis.org/measures/domainframework1 28 August 2015). Many physical activities involve skilled use of the hands. Patients with two of the most common rheumatic conditions, osteoarthritis (OA)[1] and rheumatoid arthritis (RA),[2] frequently experience affliction of the joints of the hands and surrounding tissues consisting of pain and swelling caused by underlying pathology/disease activity. As a consequence, hand function is prone to deterioration, causing decreased physical function when patients perform activities requiring their hands.[3] Therefore, physical function is both one of the core domains to be measured and reported in all clinical trials in hand OA[4 5] and a recommended core outcome for RA.[6]

At the Outcome Measures in Rheumatology (OMERACT) meeting in May 2014, a

preliminary core outcome set for hand OA was proposed, which included physical function.[5] From discussions at the meeting, it was clear that consensus is needed about what constitutes the most appropriate patient-reported outcome measurement (PROM) instrument for measuring physical function in patients with rheumatic hand conditions.[7 8] Both disease-specific and generic instruments were discussed; of the potential instruments mentioned, the Australian Canadian Osteoarthritis Hand Index (AUSCAN[9]) and the Functional Index for Hand Osteoarthritis (FIHOA[10]) were considered to be of best quality. However, both these instruments fail to fully comply with the OMERACT filter V.2.0,[11] which comprises the quality criteria discrimination, truth and feasibility. The AUSCAN does not comply with the feasibility criterion because it is copyrighted, and the FIHOA does not comply with the truth criterion (validity) because patients present at the OMERACT meeting considered it outdated. The Disabilities of the Arm, Shoulder, and Hand (DASH[12]), the Michigan Hand Questionnaire (MHQ[13]), and the Patient-Rated Wrist Hand Evaluation (PRWHE[14]) were recommended for further testing of psychometric qualities.[5] This recommendation is supported by a review concluding that existing instruments measuring physical function in patients with rheumatic hand conditions lack psychometric testing and vary in feasibility and their psychometric qualities, that is, validity, reliability, responsiveness and interpretability.[15]

Owing to the limitations of the instruments reviewed above, the Patient Reported Outcomes Measurement Information System (PROMIS) Physical Function (PF) item bank was suggested as a potential alternative (see Competing interests statement). The PROMIS PF item bank was developed using item response theory (IRT) methods[16] and can be administered in the form of different static questionnaires (short forms) or through Computer Adaptive Testing (CAT), where, after a fixed starting question, the computer selects the next question (s) based on the answer(s) to the previous question(s) until a reliable score is achieved. Although the PROMIS PF item bank is a generic instrument to assess physical function, two studies[17 18] found that the upper extremity part of the PROMIS PF CAT instrument compared favourably with the body region-specific PROM instrument DASH in orthopaedic outpatient populations. Previously, new items to assess floor and ceiling physical function have been developed in order to supplement the existing PROMIS PF item bank.[19 20]

Using a PROMIS CAT would yield several advantages to conventional questionnaires, the most important ones being comparability of scores across populations (because all PROMIS instruments are scored on a common metric),[16] and a decreased administrative burden for patients; typically with CAT, only 5–7 items need to be completed to obtain a reliable score.[21] To optimise the content validity of the PROMIS PF item bank for patients with rheumatic hand conditions, our first aim is to select those items from the PROMIS PF item bank that are relevant for patients with rheumatic hand conditions, select additional items from other existing PROMs, and develop new items if patients deem them necessary. If any additional or new items are needed, our second aim is to examine whether it is valid to score these items on the same scale as the original items from the PROMIS PF item bank. If these items can be scored on the common PROMIS PF metric, they can be added to the PROMIS PF item bank. The total set of PF items relevant for patients with hand problems can then be used as a hand-specific PF short form or a hand-specific PF CAT. Scores on these instruments will be comparable to scores from any (patient) population that completed any of the other PROMIS PF item subsets. If the new hand items cannot be scored on the common PROMIS PF metric, the total set of PF items can inform the development of a new PROM.

To measure physical function, it seems reasonable to develop an instrument that is suitable for patients across different types of rheumatic hand conditions, as patients with hand OA and RA experience similar functional problems. This goal becomes evident when comparing the brief International Classification of Functioning, Disability and Health (ICF) Core Sets for Hand Conditions,[22] OA[23] and RA,[24] which share the following domains: (1) sensation of pain, (2) mobility of joint functions, (3) muscle power functions, (4) structure of upper extremity and (5) hand and arm use. However, it is not the intention of this project to cover all of the above. The domain of the outcome measurement instrument is limited to function in order to adhere to the assumption of unidimensionality. Such an instrument may also be suitable for patients with other hand conditions, although this possibility has to be determined in future studies.

OBJECTIVES

We aim to develop a PROM to assess hand-related physical function in rheumatic hand conditions.

METHODS

The target population consists of adult patients with rheumatic conditions (defined as OA according to the American College of Rheumatology (ACR) criteria[25] or as RA according to the ACR/European League Against Rheumatism (EULAR) criteria)[26] affecting their hands. The instrument should be useful for discriminative and evaluative purposes. It was decided to focus only on these conditions, because they are the most common diagnoses with hand involvement within rheumatology.[27]
Specific steps of this study are to:

1. Perform a systematic search to identify existing PROMs to rheumatic hand conditions, and select items relevant for hand-related physical function. Selection will be based on consensus among reviewers.

2. Select items from the PROMIS PF item bank relevant to patients with rheumatic hand conditions. Selection will be based on consensus among reviewers.
3. Ensure the content validity of the selected items by discussing the items in focus groups with patients with hand OA or RA.
4. Develop new items in case of content under-representation. The items will be based on patients' input.
5. Pilot test the selected items to evaluate the comprehensibility and comprehensiveness of the items.
6. Calibrate and validate any additional and new items to examine whether it is valid to score them on the same scale as the original items from the PROMIS PF item bank.

Identifying relevant items

Relevant items will be selected from the PROMIS PF item bank by two reviewers, working independently, who have clinical experience in addressing physical function in patients with rheumatic hand conditions. Relevant items are defined as items concerning hand-related physical function. The final selection will be based on consensus between the two reviewers (or with help from a third reviewer if necessary). The item bank has already been translated into Danish, German and Dutch-Flemish (using standardised methodology and approved by the PROMIS Statistical Center).

A systematic search will be performed in PubMed, from database inception to identify PROMs for patients with hand problems. The search strategy from PubMed is shown in table 1. The website http://www.rehabmeasures.org will be searched for relevant PROMs.

Abstracts will be screened by one reviewer (LK), and the names of relevant PROMs will be extracted. Full copies of relevant PROMs will be obtained. Items relevant to measuring physical function in patients with rheumatic hand conditions will be extracted independently by two reviewers with experience in addressing physical function in patients with rheumatic hand conditions, and added to the list of items selected from the PROMIS PF item bank. Duplicates will be removed from the list by the two reviewers together; in case of a duplicate with a PROMIS item, the PROMIS item will be retained. If necessary, items will be translated by the authors into Danish, German and Dutch-Flemish. A formal forward–backward translation procedure is not considered necessary at this stage of the project.

Content validity of the item pool

To develop a relevant and comprehensive instrument, we will take a qualitative approach involving the target population in focus groups to validate the item pool developed from the PROMIS PF item bank and from the literature, and expand it if necessary.

A purposive sample of adults with hand OA and RA will be recruited through participating clinics and hospitals: Rheumatology, Ghent University Hospital and University Hospitals Leuven, Belgium; the Department of Rheumatology, Leiden University Medical Center, Leiden, the Netherlands; Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway; and the OA outpatient clinic at the Department of Rheumatology at Bispebjerg and Frederiksberg Hospital, Denmark.

Patients eligible to participate must have a clinical diagnosis of OA according to the ACR criteria[25] or RA according to the ACR/EULAR criteria,[26] affecting their hands. Patient with hand problems related to non-rheumatic disorders will not be considered eligible for inclusion. The aim is to include a variety of patients in terms of age, gender (distribution similar to the background population), diagnosis, disease severity (patient reported through the question: How severe is your disease, on a scale from 0 to 10 where 0 is not severe at all and 10 is the most severe you can imagine?), disease duration, need of aids, current pain on the Numerical Pain Rating Scale, education, ethnicity and occupation. Patients who are unable to speak or write the language at issue, or who do not understand information about the project, are not eligible. Patients will receive oral and written information prior to participating and will provide signed, informed consent before enrolling in the project.

To involve patients in identifying relevant issues, focus groups will be conducted separately at study sites in four countries (Belgium, the Netherlands, Norway and Denmark), so patients can use their native language. We anticipate conducting one focus group at each study site, each with 6–8 participants, yielding a total of three focus groups with 24–32 participants. Each focus group will be of a planned duration of 4–5 hours, including breaks with refreshments. If differences are found among the focus groups, additional focus groups will be conducted at one of the sites until saturation has been reached. At each focus group, one skilled or trained moderator and one assistant will be present.

During the focus groups, the following steps are performed, adapted from the concept mapping approach:[29] (1) participants are asked to identify "…all the things that your hand condition limits you in doing" individually, generating statements; (2) a common list of statements for the group is made, consisting of input from all participants; (3) the participants are presented with the item pool extracted from the PROMIS PF bank and other existing PROMs; (4) the group will discuss items that are included in the item pool but not mentioned by patients, as well as statements mentioned by patients who are not included in the item pool; a final item pool will be developed; (5) each participant will rate the relevance of all items on the list, on a five-point scale: 1: 'not relevant to my condition at all'; 2: 'rarely relevant to my condition'; 3: 'sometimes relevant with my condition'; 4: 'very relevant with my condition' and 5: 'essential (it would not be my condition without it)'. The mean and median ratings of statement importance are calculated for all items in order to produce a rank order

Table 1 Search strategy

Search	Filter	Query
#6	Exclusion filter	#5 NOT ('addresses'[Publication Type] OR 'biography'[Publication Type] OR 'case reports'[Publication Type] OR 'comment'[Publication Type] OR 'directory'[Publication Type] OR 'editorial'[Publication Type] OR 'festschrift'[Publication Type] OR 'interview'[Publication Type] OR 'lectures'[Publication Type] OR 'legal cases'[Publication Type] OR 'legislation'[Publication Type] OR 'letter'[Publication Type] OR 'news'[Publication Type] OR 'newspaper article'[Publication Type] OR 'patient education handout'[Publication Type] OR 'popular works'[Publication Type] OR 'congresses'[Publication Type] OR 'consensus development conference'[Publication Type] OR 'consensus development conference, nih'[Publication Type] OR 'practice guideline'[Publication Type]) NOT ('animals'[MeSH Terms] NOT 'humans'[MeSH Terms])
#5		#1 AND #2 AND #3 AND #4
#4	Function	Function*[tiab] OR *function[tiab] OR activit*[tiab] OR disability[tiab]
#3	PROM filter	Search ((HR-PRO[tiab] OR HRPRO[tiab] OR HRQL[tiab] OR HRQoL[tiab] OR QL[tiab] OR QoL[tiab] OR quality of life[tw] OR life quality[tw] OR health index*[tiab] OR health indices[tiab] OR health profile*[tiab] OR health status[tw] OR ((patient [tiab] OR self[tiab] OR child[tiab] OR parent[tiab] OR carer[tiab] OR proxy[tiab]) AND ((report[tiab] OR reported[tiab] OR reporting[tiab]) OR (rated[tiab] OR rating[tiab] OR ratings[tiab]) OR based[tiab] OR (assessed[tiab] OR assessment[tiab] OR assessments[tiab]))) OR ((disability[tiab] OR function[tiab] OR functional[tiab] OR functions[tiab] OR subjective[tiab] OR utility[tiab] OR utilities[tiab] OR wellbeing[tiab] OR well being[tiab]) AND (index[tiab] OR indices[tiab] OR instrument [tiab] OR instruments[tiab] OR measure[tiab] OR measures[tiab] OR questionnaire[tiab] OR questionnaires[tiab] OR profile [tiab] OR profiles[tiab] OR scale[tiab] OR scales[tiab] OR score[tiab] OR scores[tiab] OR status[tiab] OR survey[tiab] OR surveys[tiab])))))
#2	Precise search filter on measurement properties[28]	Search instrumentation[sh] OR 'Validation Studies'[pt] OR 'reproducibility of results'[MeSH Terms] OR reproducib*[tiab] OR 'psychometrics'[MeSH] OR psychometr*[tiab] OR clinimetr*[tiab] OR clinometr*[tiab] OR 'observer variation'[MeSH] OR 'observer variation'[tiab] OR 'discriminant analysis'[MeSH] OR reliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR 'internal consistency'[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR 'item correlation'[tiab] OR 'item correlations'[tiab] OR 'item selection'[tiab] OR 'item selections'[tiab] OR 'item reduction'[tiab] OR 'item reductions'[tiab] OR agreement[tiab] OR precision[tiab] OR imprecision[tiab] OR 'precise values'[tiab] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay [tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant [tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tiab] OR ((replicab*[tiab] OR repeated[tiab]) AND (measure[tiab] OR measures[tiab] OR findings[tiab] OR result[tiab] OR results[tiab] OR test[tiab] OR tests[tiab])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR 'known group'[tiab] OR 'factor analysis'[tiab] OR 'factor analyses'[tiab] OR dimensionality[tiab] OR subscale*[tiab] OR 'multitrait scaling analysis'[tiab] OR 'multitrait scaling analyses'[tiab] OR 'item discriminant'[tiab] OR 'interscale correlation'[tiab] OR 'interscale correlations'[tiab] OR ((error[tiab] OR errors[tiab]) AND (measure*[tiab] OR correlat*[tiab] OR evaluat*[tiab] OR accuracy [tiab] OR accurate[tiab] OR precision[tiab] OR mean[tiab])) OR 'individual variability'[tiab] OR 'variability analysis'[tiab] OR

Continued

Table 1 Continued

Search	Filter	Query
		(uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR 'standard error of measurement'[tiab] OR sensitiv* [tiab] OR responsive*[tiab] OR (small*[tiab] AND (real[tiab] OR detectable[tiab])) AND (change[tiab] OR difference[tiab])) OR 'meaningful change'[tiab] OR 'minimal important change'[tiab] OR 'minimal important difference'[tiab] OR 'minimally important change'[tiab] OR 'minimally important difference'[tiab] OR 'minimal detectable change'[tiab] OR 'minimal detectable difference'[tiab] OR 'minimally detectable change'[tiab] OR 'minimally detectable difference'[tiab] OR 'minimal real change'[tiab] OR 'minimal real difference'[tiab] OR 'minimally real change'[tiab] OR 'minimally real difference'[tiab] OR 'ceiling effect'[tiab] OR 'floor effect'[tiab] OR 'item response model'[tiab] OR IRT[tiab] OR Rasch[tiab] OR 'Differential item functioning'[tiab] OR DIF[tiab] OR 'computer adaptive testing'[tiab] OR 'item bank '[tiab] OR 'cross-cultural equivalence'[tiab]
#1	Hand	Search hand[tiab]

PROM, patient-reported outcome measurement.

to guide future selection of items for instrument validation, that is, patients' ratings of item relevance will be considered in case the results of the IRT analyses suggest that two or more items cover the same level of difficulty.

Further, participants will be asked to provide demographic data including age, gender, diagnosis, disease severity, disease duration, need of aids, current pain on the Numerical Pain Rating Scale, education, occupation and ethnicity in order to describe our study population and estimate the transferability of the results. Further, self-reports on height, weight, subset of disease (which areas are affected), comorbidities and affection of other joints are gathered.

Transformation to items and translation

For each new statement from the final item pool, an item is formulated in English. The newly generated items are framed to fit the existing items from the PROMIS PF item bank (eg, "Are y*ou able to…*") Likewise, the items selected from existing PROMs are reframed to fit the existing items from the PROMIS PF item bank. The newly framed items and the reframed items from the final item pool are then (back) translated into German, Dutch-Flemish, Norwegian and Danish using standardised methodology approved by the PROMIS Statistical Center. This process is guided by the original statements produced in the local focus groups when possible. The final item pool constitutes the tentative item bank (in four languages: English, German, Dutch-Flemish and Danish) for hand-related physical function for rheumatic conditions involving hands.

PILOT TESTING

Pilot testing of the tentative item bank for hand-related physical function is performed to establish comprehensibility and comprehensiveness of the items. Participants will be recruited from the involved study sites. At each site, a stratified sample of about 10 patients with a rheumatic hand condition will be recruited, varying in terms of age, gender, diagnosis, disease severity, disease duration, need of aids, current pain on the Numerical Pain Rating Scale, education and occupation.

A qualitative cognitive interview method will be used, such as a think aloud method, Three Step Test interviews, or cognitive debriefing to evaluate the comprehensibility and comprehensiveness of the items. Skilled or trained interviewers will be used. The interviews will be recorded, and two persons (LK and a local researcher) will be involved in the analyses. Suggested modifications to the items will be discussed in the project group. If major changes are made, a new pilot test will be performed.

CALIBRATION AND VALIDATION OF THE ITEM BANK

In each participating country, we intend to recruit a sample of 200 patients (depending on funding) to obtain a sample of at least 800 patients in total, following

the recommendations of the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group which suggested 'a minimum of 200 respondents per group as a requirement for logistic regression Differential Item Functioning (DIF) analyses'.[30] Since we are aiming at investigating language DIF using ordinal logistic regression, n=200 per country is a reasonable number. Patients eligible for participation must have a clinical diagnosis of OA according to the ACR criteria[25] or RA according to the ACR/EULAR criteria[26] affecting their hands. Patients who are unable to speak or write the language at issue, or who do not understand information about the project, are not eligible. Patients will receive oral and written information prior to participating and will provide signed, informed consent before enrolling in the project. For participating countries where Ethics Committee approval is required for this type of project, approval will be obtained before project start.

Patients will be invited by email or letter to fill in a web-based (digital) or paper-and-pencil (paper) questionnaire. Both options are available as it has been shown that the results from paper administration and electronic administration of questionnaires are comparable.[31–34] For the digital questionnaire, patients will receive personal login codes. The questionnaire will include the full item bank developed under objectives 1–5. Participants will be asked to provide demographic data including age, gender, diagnosis, disease severity, disease duration, need of aids, education, occupation and the PROMIS Global Health item bank (further 10 items, including an item on current pain on the Numerical Pain Rating Scale), in order to describe our study population and estimate the transferability of the results. Two further PROMs, the FIHOA and the DASH, were selected for evaluating construct validity based on evidence of their measurement properties from existing systematic reviews of PROMs.

Calibration analyses will be conducted at the Department of Psychosomatic Medicine at Charité—Universitätsmedizin Berlin, Germany. The analyses will follow the methods used for calibrating the original PROMIS PF item bank in US samples,[35] which were largely based on the general PROMIS approach.[36] Unidimensionality will be examined using confirmatory factor analyses (CFA). Since previous studies suggested a potential two-factor solution,[35] the assumed unidimensional factor solution will be compared with an alternative two factorial model to investigate if the new items measuring hand function can be scaled on the same metric as the remaining items of the entire PF item bank. The two-factor solution is defined as hand-related and finger-related activities (fine motor actions), as opposed to the rest of the item bank (gross motor actions). Preliminary model fit will be evaluated based on commonly applied fit indices; items with factor loadings <0.60 will be considered candidates for exclusion. To evaluate model fit, the following fit indices will be examined: Comparative Fit Index (CFI), Tucker Lewis Index (TLI), Root Means Square Error of Approximation (RMSEA) and Standardised Root Mean Residuals (SRMR). The criteria for unidimensionality include CFI>0.95, TLI>0.95, RMSEA<0.06 and SRMR<0.08.[37 38] However, since common fit statistics may be misleading in dimensionality assessment of patient-reported outcomes (PROs),[39] exploratory-based techniques, such as explained common variance (ECV), were recently suggested as being most appropriate to test for unidimensionality.[40] Therefore, we will estimate an additional exploratory bifactor model to determine the ECV as an indicator for unidimensionality (ECV>0.60).[40] Local dependence will be evaluated by examining the residual correlation matrix resulting from the single factor CFA; a residual correlation of a pair of items >0.25 will indicate local dependence. In such cases, the item showing a higher accumulated residual correlation with the remaining items will be considered for exclusion.

Monotonicity of the items will be evaluated by fitting a non-parametric IRT model using Mokken scaling[36] and visual inspection of resulting item characteristic curves. After IRT assumptions are evaluated, including the aforementioned monotonicity check, a Graded Response Model (GRM) will be fitted to the data to calibrate the item parameters.[41] To assess the fit of the GRM and the degree to which possible misfit affects the IRT model, S-X2 statistic will be used. Items that show S-X2 p values <0.001 will indicate a poor fit. DIF for age (median split), gender and language will be evaluated using ordinal logistic regression models as described by Nagelkerke.[42] An R2 change of ΔR2>0.03 will indicate noticeable DIF. Based on the GRM, item information functions will be calculated for each item, which will determine the specific contribution of an item to the precision of the entire item bank on a given level of physical function. Summarising information functions (IIFs) will enable estimation of the SEs (as a parameter of reliability) across the score range of the entire item bank. If DIF is not indicated for English language (USA) versus other languages (European countries), IRT θ scores will be transformed into T-scores anchored on the US item parameters of the PROMIS PF item bank. Construct validity will be evaluated by correlating the T-scores of the item bank to the scores on the legacy instruments. A correlation of >0.50 is expected.[43]

ETHICS AND DISSEMINATION

PROs are frequently defined as 'any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else'. We anticipate that the development of a PROM to assess hand-related physical function in rheumatic hand conditions will provide researchers and clinicians with an efficient, reliable and valid assessment tool. Our ultimate goal will be

to develop a PROM for use in clinical trials where PRO end points are used to support product labelling claims —an instrument that can subsequently be applied in clinical practice as well. The results of this global initiative will be discussed and thus disseminated at EULAR, Osteoarthritis Research Society International (OARSI), ACR and hopefully be presented as a workshop at the next OMERACT meeting (2018).

Limitations of this project could be:

- The domain of the outcome measurement instrument is limited to function, based on the assumption that function is a pivotal construct that also relates to other aspects of a health condition (eg, pain).
- The patient population is limited to OA and RA with hand involvement.
- The risk that the newly developed hand-related items are not measuring the same (unidimensional) PF construct as defined by the original PROMIS PF construct. Consequently, if this is the case, we will not be able to calibrate these items on the PROMIS metric.

The methodology that this project is based on represents the current best practice[44] within development of a patient-reported questionnaire of hand physical function for use in common rheumatic conditions. It is anticipated that there is a clinical need as there is no consensus about the best choice of hand PROM; for example, AUSCAN[9] and FIHOA,[10] both developed for hand OA, are the most frequently used and respected, but the former is associated with licence use, restricting its feasibility and widespread application, and some of the items in the latter have been criticised lately for not being contemporary. For RA, Michigan hand questionnaire[13] and DASH[12] are more frequently used, but need further psychometric evaluation.

Other advantages to the approach taken in this project will be the linking to the existing PROMIS programme and using the CAT technology where only a subset of items are needed compared with paper form PROMs. This potentially reduces the burden on patients.

This study will be carried out in accordance with the Declaration of Helsinki. Ethics approvals will be obtained where necessary. Signed informed consent will be obtained from all participants. We aim to disseminate the results of the study through publication in international peer-reviewed journals and at international conferences.

Author affiliations
[1]The Parker Institute, Bispebjerg & Frederiksberg Hospital, Copenhagen, Denmark
[2]Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands
[3]The EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands
[4]Department of Public Health, University of Southern Denmark, Odense, Denmark
[5]Department of Physical and Occupational Therapy, Copenhagen University Hospital Bispebjerg Frederiksberg, Copenhagen, Denmark
[6]Department of Psychosomatic Medicine, Center of Internal Medicine and Dermatology, Charité—Universitätsmedizin Berlin, Berlin, Germany
[7]Department of Public Health Innovation, Population Health Strategic Research Centre, School of Health and Social Development, Deakin University, Burwood, Victoria, Australia
[8]Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands
[9]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands
[10]Department of Development and Regeneration KU Leuven, Rheumatology, Skeletal Biology and Engineering Research Center, University Hospitals Leuven, Leuven, Belgium
[11]Department of Rheumatology, Ghent University Hospital, Ghent, Belgium
[12]National Advisory Unit on Rehabilitation in Rheumatology, Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway
[13]Department of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway
[14]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

Acknowledgements The authors wish to acknowledge the contributions of the staff members of the participating study sites, and the patients participating in the focus group interviews.

Contributors LK is the principal investigator for this study. LK, CBT and RC are co-responsible for all phases of this study. LK, CBT, RC, EEW, MH, SN, GL, MK, RWe, RWi, IK, IKH, BS, RG and HB participated in the design of the study and drafting of the protocol. All authors have given final approval for the protocol to be published.

Funding The study was indirectly supported by unrestricted grants provided to the Parker Institute from The Oak Foundation.

Competing interests CBT and RWe are coordinators of the Dutch-Flemish PROMIS group (http://www.dutchflemishpromis.nl). CBT, SN and GL are part of the PROMIS International Group. GL is doing his PhD on the PROMIS PF Item Bank.

Patient consent Obtained.

Ethics approval Commissie Medische Ethiek—toetsingscommissie UZ Leuven Campus Gasthuisberg Herestraat 49 B-3000 Leuven Belgium.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Cross M, Smith E, Hoy D, *et al*. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. *Ann Rheum Dis* 2014;73:1323–30.
2. Cross M, Smith E, Hoy D, *et al*. The global burden of rheumatoid arthritis: estimates from the global burden of disease 2010 study. *Ann Rheum Dis* 2014;73:1316–22.
3. March L, Smith EU, Hoy DG, *et al*. Burden of disability due to musculoskeletal (MSK) disorders. *Best Pract Res Clin Rheumatol* 2014;28:353–66.
4. Bellamy N, Kirwan J, Boers M, *et al*. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. *J Rheumatol* 1997;24:799–802.
5. Kloppenburg M, Boyesen P, Visser AW, *et al*. Report from the OMERACT Hand Osteoarthritis Working Group: set of core domains and preliminary set of instruments for use in clinical trials and observational studies. *J Rheumatol* 2015;42:2190–7.
6. Felson DT, Anderson JJ, Boers M, *et al*. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. *Arthritis Rheum* 1993;36:729–40.

7. Kloppenburg M, Boyesen P, Smeets W, *et al*. Report from the OMERACT Hand Osteoarthritis Special Interest Group: advances and future research priorities. *J Rheumatol* 2014;41:810–18.
8. Visser AW, Bøyesen P, Haugen IK, *et al*. Instruments measuring pain, physical function or patient global assessment in hand osteoarthritis. *J Rheumatol* 2015;42:2118–34.
9. Bellamy N, Campbell J, Haraoui B, *et al*. Dimensionality and clinical importance of pain and disability in hand osteoarthritis: development of the Australian/Canadian (AUSCAN) Osteoarthritis Hand Index. *Osteoarthr Cartil* 2002;10:855–62.
10. Dreiser RL, Maheu E, Guillou GB, *et al*. Validation of an algofunctional index for osteoarthritis of the hand. *Rev Rhum Engl Ed* 1995;62(Suppl 1):43S–53S.
11. Boers M, Kirwan JR, Wells G, *et al*. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. *J Clin Epidemiol* 2014;67:745–53.
12. Beaton DE, Katz JN, Fossel AH, *et al*. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. *J Hand Ther* 2001;14:128–46.
13. Chung KC, Pillsbury MS, Walters MR, *et al*. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. *J Hand Surg Am* 1998;23:575–87.
14. MacDermid JC, Turgeon T, Richards RS, *et al*. Patient rating of wrist pain and disability: a reliable and valid measurement tool. *J Orthop Trauma* 1998;12:577–86.
15. Poole JL. Measures of hand function: Arthritis Hand Function Test (AHFT), Australian Canadian Osteoarthritis Hand Index (AUSCAN), Cochin Hand Function Scale, Functional Index for Hand Osteoarthritis (FIHOA), Grip Ability Test (GAT), Jebsen Hand Function Test (JHFT), and Michigan Hand Outcomes Questionnaire (MHQ). *Arthritis Care Res (Hoboken)* 2011;63(Suppl 11):S189–99.
16. Rose M, Bjorner JB, Gandek B, *et al*. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. *J Clin Epidemiol* 2014;67:516–26.
17. Doring AC, Nota SP, Hageman MG, *et al*. Measurement of upper extremity disability using the Patient-Reported Outcomes Measurement Information System. *J Hand Surg Am* 2014;39:1160–5.
18. Tyser AR, Beckmann J, Franklin JD, *et al*. Evaluation of the PROMIS Physical Function Computer Adaptive Test in the upper extremity. *J Hand Surg Am* 2014;39:2047–51.
19. Bruce B, Fries J, Lingala B, *et al*. Development and assessment of floor and ceiling items for the PROMIS physical function item bank. *Arthritis Res Ther* 2013;15:R144.
20. Fries JF, Lingala B, Siemons L, *et al*. Extending the floor and the ceiling for assessment of physical function. *Arthritis Res Ther* 2014;66:1378–87.
21. Fries JF, Witter J, Rose M, *et al*. Item response theory, computerized adaptive testing, and PROMIS: assessment of physical function. *J Rheumatol* 2014;41:153–8.
22. Rudolf KD, Kus S, Chung KC, *et al*. Development of the International Classification of Functioning, Disability and Health core sets for hand conditions—results of the World Health Organization International Consensus process. *Disabil Rehabil* 2012;34:681–93.
23. Dreinhofer K, Stucki G, Ewert T, *et al*. ICF Core Sets for osteoarthritis. *J Rehabil Med* 2004;(44 Suppl):75–80.
24. Stucki G, Cieza A, Geyh S, *et al*. ICF Core Sets for rheumatoid arthritis. *J Rehabil Med* 2004;(44 Suppl):87–93.
25. Altman R, Alarcon G, Appelrouth D, *et al*. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hand. *Arthritis Rheum* 1990;33:1601–10.
26. Aletaha D, Neogi T, Silman AJ, *et al*. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/ European League Against Rheumatism collaborative initiative. *Ann Rheum Dis* 2010;69:1580–8.
27. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;386:743–800.
28. Terwee CB, Jansma EP, Riphagen II, *et al*. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;18:1115–23.
29. Trochim W, Kane M. Concept mapping: an introduction to structured conceptualization in health care. *Int J Qual Health Care* 2005;17:187–91.
30. Scott NW, Fayers PM, Aaronson NK, *et al*. Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. *Health Qual Life Outcomes* 2010;8:81.
31. Waehrens EE, Bliddal H, Danneskiold-Samsoe B, *et al*. Differences between questionnaire- and interview-based measures of activities of daily living (ADL) ability and their association with observed ADL ability in women with rheumatoid arthritis, knee osteoarthritis, and fibromyalgia. *Scand J Rheumatol* 2012;41:95–102.
32. Muehlhausen W, Doll H, Quadri N, *et al*. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. *Health Qual Life Outcomes* 2015;13:167.
33. Bjorner JB, Rose M, Gandek B, *et al*. Method of administration of PROMIS scales did not significantly impact score level, reliability, or validity. *J Clin Epidemiol* 2014;67:108–13.
34. Bjorner JB, Rose M, Gandek B, *et al*. Difference in method of administration did not significantly impact item response: an IRT-based analysis from the Patient-Reported Outcomes Measurement Information System (PROMIS) initiative. *Qual Life Res* 2014;23:217–27.
35. Rose M, Bjorner JB, Becker J, *et al*. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol* 2008;61:17–33.
36. Reeve BB, Hays RD, Bjorner JB, *et al*. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care* 2007;45(Suppl 1):S22–31.
37. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equat Model* 1999;6:1–55.
38. Dworkin RH, Turk DC, Peirce-Sandner S, *et al*. Research design considerations for confirmatory chronic pain clinical trials: IMMPACT recommendations. *Pain* 2010;149:177–93.
39. Fayers PM, Aaronson NK. “It ain’t over till the fat lady sings”: a response to Cameron N. McIntosh, improving the evaluation of model fit in confirmatory factor analysis. *Qual Life Res* 2012;21:1623–4.
40. Reise SP, Scheines R, Widaman KF, *et al*. Multidimensionality and structural coefficient bias in structural equation modeling: a bifactor perspective. *Educ Psychol Meas* 2013;73:5–26.
41. van der Linden WJ, Hambleton RK. *Handbook of modern item response theory*. Berlin, Germany: Springer, 1997.
42. Nagelkerke NJD. A note on general definition of the coefficient of determination. *Biometrika* 1991;78:691–2.
43. Prinsen CAC, Vohra S, Rose MR, *et al*. How to select outcome measurement instruments for outcomes included in a ‘Core Outcome Set’—a practical guideline. *Trials* 2016;17:449.
44. de Vet HC, Terwee CB, Mokkink LB, *et al*. *Measurement in medicine a practical guide*. Cambridge University Press, 2011.